3А

УДК 004.89:004.93

Д.С. Азаренко
Институт проблем искусственного интеллекта
МОН Украины и НАН Украины, г. Донецк
Украина, 83048, г. Донецк, ул. Артема, 118-б

Детектирование нот на изображении нотного текста на основе результатов его структурного анализа

D.S. Azarenko
Institute of Artificial Intelligence
MES of Ukraine and NAS of Ukraine, c. Donetsk
Ukraine, 83048, c. Donetsk, Artema st., 118 b

Detection of Scores to the Note Text Image Based on its Structural Analysis

Д.С. Азаренко
Інститут проблем штучного інтелекту
МОН України і НАН України, м. Донецьк
Україна, 83048, м. Донецьк, вул. Артема 118-б

Детектування нот на зображенні нотного тексту на основі результатів його структурного аналізу

В статье предложен подход к структурному анализу нотного текста, целью которого является детектирование значимых элементов изображения. Для отделения фона от значимых объектов разработана процедура предобработки изображения, использующая порог яркости. Структурный анализ проводится с учетов иерархической структуры нотного текста, для поиска элементов которой применяется преобразование Хафа. Детектирование нот проводится с использованием знания геометрии искомых объектов с помощью эллиптической маски.
Ключевые слова: структурирование изображения, гистограмма яркости, преобразование Хафа, эллиптическая маска.

In the paper, the approach to the structural analysis of music format aimed at the detection of significant elements of the image is proposed. To separate the background from the important objects, procedure of image preprocess, which uses the brightness threshold, is developed. The structural analysis is performed taking into account the hierarchical structure of the music format, to search for items that uses the Hough transform. Detection of music is performed with the required knowledge of geometry objects using elliptical mask.
Key Words: image structuring, brightness histogram, Hough transform, elliptical mask.

У статті запропоновано підхід до структурному аналізу нотного тексту, метою якого є детектування значущих елементів зображення. Для відділення фону від значимих об’єктів розроблена процедура передобробки зображення, що використовує поріг яскравості. Структурний аналіз проводиться з обліків ієрархічної структури нотного тексту, для пошуку елементів якої застосовується перетворення Хафа. Детектування нот проводиться з використанням знання геометрії шуканих об’єктів за допомогою еліптичної маски.
Ключові слова: структурування зображення, гістограма яскравості, перетворення Хафа, еліптична маска.

Введение

Первое печатное издание нотного текста появилось в 16 веке, и за этот период накопилось огромное количество произведений, составляющих культурное наследие человечества. Но бумажные издания, к сожалению, обладают рядом существенных недостатков, среди которых – недолговечность, относительно сложный процесс копирования и невозможность обработки при помощи ЭВМ. Для литературных текстов существуют десятки эффективных программ, которые, используя технологии оптического распознавания текста, позволяют конвертировать изображения рукописного, машинописного или печатного текста в электронный вид для автоматизации систем учета в бизнесе или для публикации текста на веб-странице.

Программы по распознаванию сканированного нотного текста находятся, можно сказать, в зачаточном состоянии, что связано с проблемами, возникающими при оптическом распознавании нотного текста. Это объясняется сложной структурой нотной графики, где могут присутствовать наряду со сложными элементами (гитарные табулатуры, музыкальные ключи, крестообразные нотные головки, символ дубль диеза, скобки у знаков альтерации, аппликатура, форшлаги, символы арпеджио, педаль и конец педали, повторы, репризы нот и тактов) секции с буквами, включая рукописные.

Яркими примерами таких программы являются «Midiscan» и «Capella-scan», которые распознают сканированный нотный текст и превращают его в стандартный MIDI-файл. Кроме того, модули распознавания нотной графики внедрены в многочисленные программы-нотаторы. После распознавания пользователь видит на экране, что было распознано из нотного текста, и может его подредактировать, перед тем, как перевести в MIDI-формат. Однако распространенным мнением об эффективности работы модулей распознавания является «быстрее набрать все руками, чем потом исправлять». Таким образом, задача качественного максимально автоматизированного оптического распознавания печатного нотного текста является актуальной и востребованной.

Данная работа выполнялась в рамках гранта «Создание информационной технологии автоматического оптического распознавания печатного нотного текста», конечным результатом выполненных по нему работ является создание информационной технологии оптического распознавания печатного нотного текста, которая позволит сделать процесс распознавания более автоматизированным и робастным по отношению к качеству исходных данных. Это подразумевает для конечного пользователя минимальное вмешательство в процесс настройки системы распознавания нотного текста и обработки результатов распознавания.

В статье речь пойдет об одном из этапов, без которого невозможно осуществить распознавание нот на изображении, о проведении структурного анализа нотного изображения и детектировании нот на нем.

Цель работы – разработка метода детектирования нот на цифровом изображении, который позволит сделать процесс распознавания более автоматизированным и робастным по отношению к качеству исходных данных. Это подразумевает для конечного пользователя минимальное вмешательство в процесс настройки системы распознавания нотного текста и обработки результатов распознавания.

Для достижения поставленной цели необходимо решить следующие задачи:

– предварительная обработка изображения, заключающаяся в его бинаризации и шумоочистке;

– выделение структурных элементов изображения;

– детектирование нот на полученных структурных элементах.

Предобработка изображения, содержащего нотный текст

Все действия по обработке, сегментации и распознаванию будут проводиться над изображением страницы нотного текста, полученным со сканера.

Нотный текст носит четко выраженный бинарный характер (наличие только значимых объектов и фона), а изображения, поступающие на обработку, могут содержать разное количество цветов. Поэтому необходим полученный порог отделения фона от значимых элементов. В дальнейшем этот порог будет использоваться в методах обработки изображения и при необходимости корректироваться для получения более качественных результатов. Данная задача решается на два этапа:

1) приведение изображения к оттенкам серого;
2) построение гистограммы и определение порога.

Приведение изображения к оттенкам серого производится путем пересчета всех цветов по формуле:

$$Y = 0.222\ Red + 0.707\ Green + 0.071\ Blue,$$

где Y – искомое значение в сером цвете; *Red* – значение красного коэффициента в палитре RGB; *Green* – значение зеленого коэффициента в палитре RGB; *Blue* – значение голубого коэффициента в палитре RGB.

Коэффициенты в данной формуле обусловлены разной чувствительностью человеческого глаза к компонентам цвета.

Для бинаризации полученного изображения необходимо построить гистограмму яркости. Для большинства изображений нотного текста она будет иметь вид, подобный представленному на рис. 1.

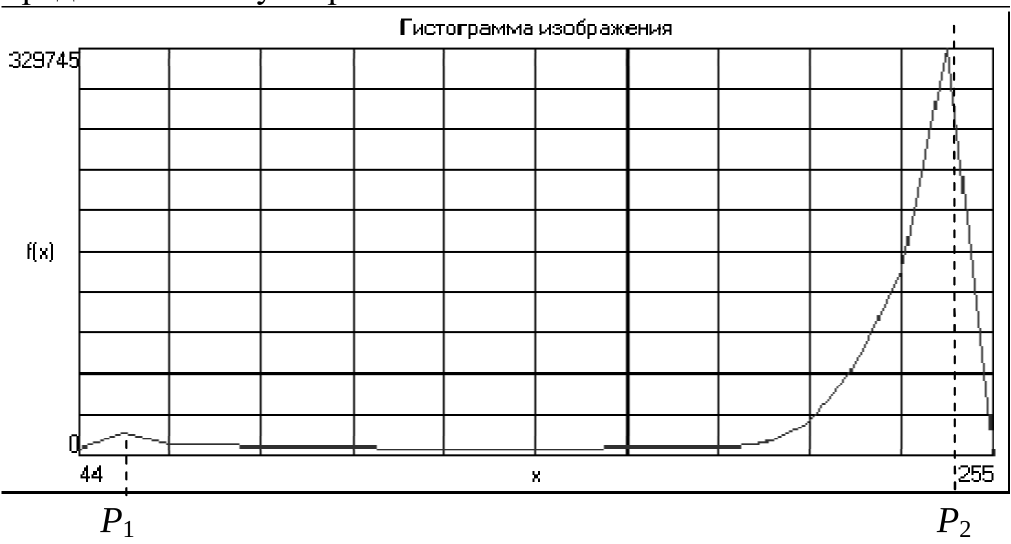


Рисунок 1 – Гистограмма яркости изображения нотного текста

Как видно из рис. 1, гистограмма содержит два пика в точках P_1 и P_2. Необходимо определить порог бинаризации так, чтобы он обеспечил наилучшее отделение фона от значимых объектов изображения. Для этого необходимо найти локальный минимум между двумя пиками. Если найденных минимумов несколько, то выбираем тот, который ближе к правому пику. Обозначим найденную точку через M.

Теперь необходимо определить момент, когда происходит резкое возрастание значений гистограммы $f(x)$ на участке между выбранным минимумом M и правым пиком. Под «резким» будем понимать возрастание значений гистограммы не менее чем в 3 раза при увеличении аргумента на 1.

Следовательно, для точки «резкого возрастания» (обозначим ее через *N*) выполняется условие:

$$\left(\frac{f(x+1)}{f(x)}<3\right)\wedge\left(\frac{f(N+1)}{f(N)}\right)\geq 3 \quad \text{при} \quad M\leq x<N$$

Величину *D* – смещения точки *N*, как показали численные исследования, смещение необходимо для более корректного вычисления порога яркости, по которому фон отделяется от значимых объектов:

$$D=1-\frac{f(P_1)}{f(P_2)} \quad .$$

Для вычисления значения искомого порога яркости используется формула, полученная экспериментально:

$$P=\frac{N+M}{2}+(N-M)\cdot D \quad .$$

Поскольку исходное изображение содержит шум, следующим шагом в предобработке является шумоочистка. Помимо шума изображение нотного текста содержит также много мелких значимых частей, которые необходимо распознать. Такими фрагментами могут являться словесный текст, обозначения акцентов, стаккато и т.д. Кроме того, в эту категорию попадают точки, означающие удлинение ноты в полтора раза. Учитывая специфику нотного текста (большинство объектов, подлежащих распознаванию, пересекаются с нотными станами), наиболее разумным действием по шумоочистке представляется выполнение сегментации и удаление всех сегментов, которые имеют относительно маленький размер. Безусловно, это приведет к удалению объектов, которые не пересекаются с линиями нотного стана. Как следствие, за шум могут быть приняты нужные части изображения. Поэтому целесообразно перед удалением объектов провести их предварительное распознавание, такое распознавание будет проводиться при необходимости в последнюю очередь как уточняющее.

Распознавание печатного нотного текста невозможно без процедуры его структурирования, поскольку выделение структурных единиц позволит упростить последующее распознавание и повысить его эффективность.

Структурный анализ изображения нотного текста

В результате выполнения структурного анализа изображения детали, способные помешать распознаванию нотного текста, убираются, но при этом информация, которую они несут, сохраняется.

Эта задача осложняется разнообразием объектов структурирования. Для каждого из них необходимо выработать свой подход и методы обработки. Данная задача для каждого объекта разбивается на последовательно выполняющиеся процедуры:

1) поиск объекта и его идентификация;
2) сохранение свойств объекта;
3) удаление объекта.

Задачи обработки и распознавания изображений требуют большого количества повторяющихся операций. Обработка изображения целиком требует больших затрат времени и ресурсов системы. В то же время известно, что нотный текст имеет иерархическую структуру (рис. 2). Если выполнить структурирование изображения и работать с каждой из структур отдельно, то это позволит уменьшить количество вычислений.

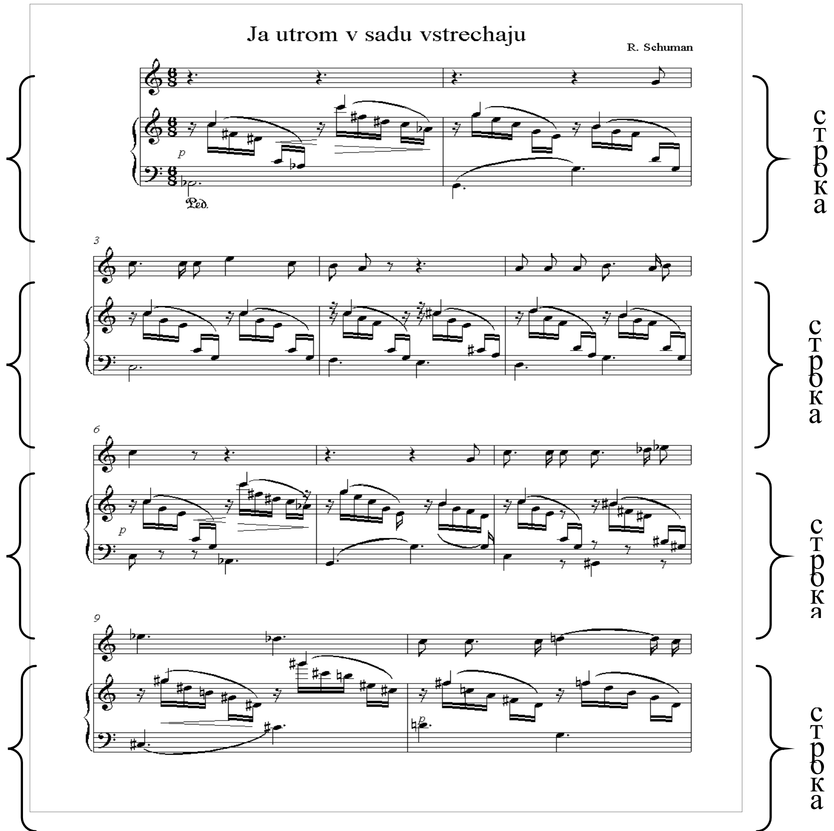


Рисунок 2 – Изображение нотного текста

Как видно из рис. 2, на листе нотного текста можно выделить следующие структурные элементы:

1) строка (набор нотоносцев, начала которых соединены общей тактовой чертой);

2) группа нотоносцев (набор нотоносцев, соединенных вертикальными отрезками, каждый из которых в пределах нотоносца дает тактовую черту);

3) нотоносец.

Каждый последующий элемент является структурной единицей предыдущего. Так, из нотоносцев составляются группы, а из групп – строки. В пределах каждой структурной единицы работа ведется с ее изображением. Это позволяет в значительной степени сократить объем вычислений и использованной памяти.

Для нахождения строк необходимо сегментировать бинаризованное и очищенное от шума изображение. Каждый полученный сегмент будет являться строкой.

Переходим к следующему уровню иерархии: каждую найденную строку делим на группы нотоносцев. Группы нотоносцев обладают следующим уникальным свойством: все нотоносцы группы соединены слева непрерывной вертикальной линией, а справа – нет. Первоначально необходимо найти эти вертикальные линии (линия выделена жирным на рис. 3), и здесь возникает очередная проблема.
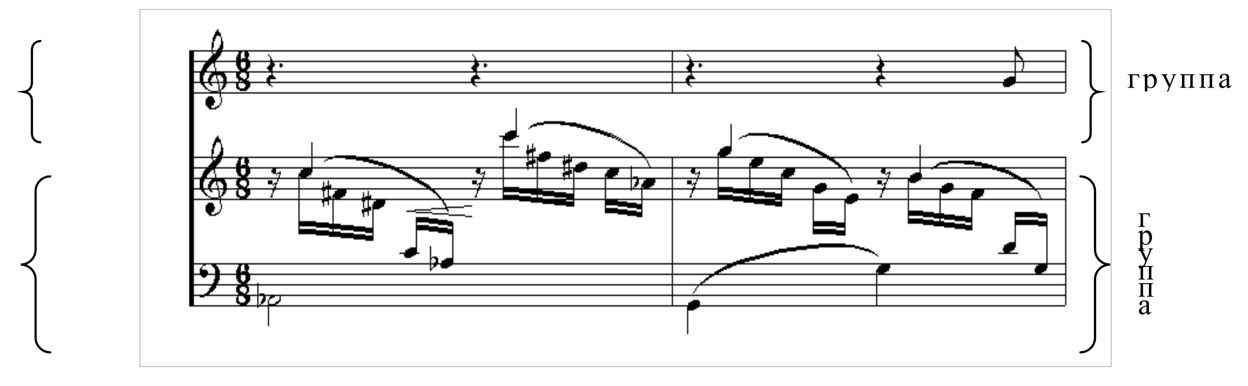


Рисунок 3 – Группы в строках

В идеальном случае угол наклона линий нотного стана нулевой. Однако эта ситуация в большинстве случаев недостижима. Кроме того, как показала практика, эти линии могут прерываться из-за ошибок во время сканирования. Учитывая это, был использован метод поиска линий, основанный на преобразовании Хафа, широко используемый системами обработки изображений для обнаружения прямых.

В простейшем случае преобразование Хафа является линейным и использует уравнение прямой с угловым коэффициентом:

$$y = mx + b,$$

где m – тангенс угла наклона прямой к оси OX, b – координата точки пересечения прямой с осью OY.

Основная идея преобразования Хафа – учесть характеристики прямой не как геометрического места точек изображения, а в терминах ее параметров, т.е. m и b. Прямая представляется в виде точки с координатами (b, m) в пространстве параметров. Для каждой точки в пространстве (x, y) в пространстве параметров (b,m) будет соответствовать прямая, чем чаще через точку в пространстве (b, m) будет проходить прямая, тем больше будет коэффициент z у этой точки. Нас интересует значение m, так как это tan(alfa) угла наклона прямой, и b как значение положения линии. А также значение z в точке (b, m), так как, чем больше это значение, тем вероятнее положения линии с данными параметрами. Зная эти параметры, мы получаем положение интересующей нас линии. В дальнейшем преобразование Хафа применяется для поиска всех интересующих нас линий на изображении нотного текста.

Для выделения групп необходимо удалить найденную вертикальную линию и повторить процедуру сегментации. Полученные сегменты будут являться группами нотоносцев.

Для завершения структурирования изображения необходимо выделить нотоносцы, входящие в каждую из найденных групп. Для этого необходимо найти все тактовые линии по описанному методу поиска линий. Найденные линии необходимо удалить (рис. 4).
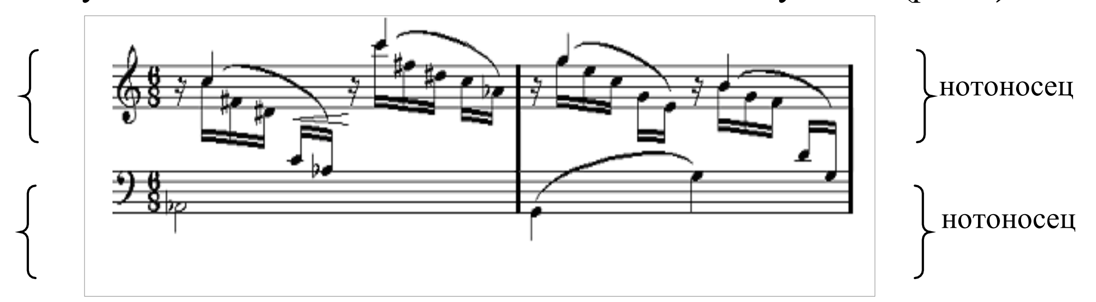


Рисунок 4 – Нотоносцы в группе

После сегментации полученного изображения каждый сегмент будет представлять собой отдельный нотоносец.

Последний шаг в процедуре структурирования изображения – удаление нотоносцев. Причем сделать это надо, не задев ноты и другие объекты распознавания.

Положение и наклон линий, а также их толщину мы получим после применения преобразования Хафа для поиска горизонтальных линий. Следует отметить, что простое удаление линий приводит к повреждению тех элементов изображения, которые в дальнейшем подлежат распознаванию. Поэтому необходимо проверять наличие объектов на линии перед удалением. Для этого производится проверка на наличие черных фрагментов с одной из сторон линии. Если они есть, то данный фрагмент должен быть больше, чем $Dl/3$, где Dl – это расстояние между линиями нотного стана. При таком подходе

возможно небольшое обрезание объектов (нот, знаков и т.д.) по краям. В то же время это делает возможным полное удаление линии (неполное удаление вызвало больше проблем при дальнейшем распознавании).

После удаления всех линий нотоносца изображение содержит объекты, подлежащие распознаванию: ноты, знаки альтерации, паузы и т.п. (рис. 5).

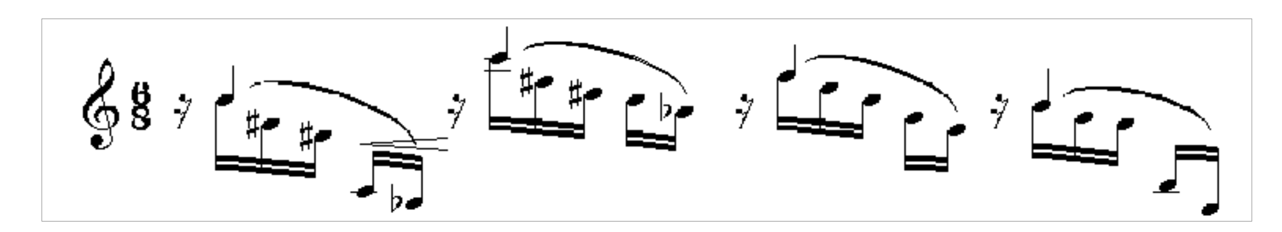


Рисунок 5 – Изображение после удаления линий нотоносца

Поскольку сохранены свойства удаленных объектов: угол наклона, масштаб (*Dl*), координаты всех горизонтальных и вертикальных линий, то при дальнейшем распознавании не составит труда восстановить принадлежность того или иного объекта к определенному нотоносцу и определить взаимное расположение уже распознанных объектов и объектов, которые были удалены.

Следующим шагом, необходимым для распознавания, является детектирование нот на изображении.

Детектирование нот с использованием знаний о геометрии искомых объектов

Достаточно точным математическим описанием ноты является эллипс. Определив для него длину большой и малой полуосей (*a*, *b*) и угол поворота относительно оси абсцисс, можно задать маску, при помощи которой и будет осуществляться дальнейший поиск нот.

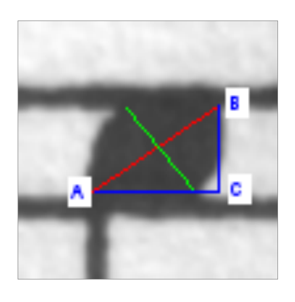


Рисунок 6 – Нота с условными обозначениями

На рис. 6 показана нота и условные обозначения, используемые в дальнейших рассуждениях. Как выяснилось в результате проведённых исследований, угол наклона эллипса, описывающего ноту, равен 37 градусам, а соотношение между полуосями b/a = 0,66. Зная расстояние между линиями нотного стана (*BC*) и угол наклона ноты, можно определить один из диаметров эллипса $AB = BC/sin(37)$, а, следовательно, и большую полуось $a = AB/2$.

Основываясь на приведенных рассуждениях, ноты на изображении нотного текста можно детектировать как эллипсы с вычисленными параметрами. В то же время необходимость учёта случайной природы дефектов обрабатываемого изображения приводит к необходимости введения допусков на значения полуосей эллипсов. При

увеличении длин полуосей количество кандидатов в ноты уменьшается до тех пор, пока ноты не перестают обнаруживаться вообще. На рис. 7 приведены результаты детектирования нот с разными радиусами эллиптических масок. Причём, чем темнее точка, обозначающая центр ноты, тем большее значение радиусов соответствует ей.



Рисунок 7 – Результат детектирования нот

Очевидно, что в данном случае наибольший интерес представляют точки, соответствующие эллипсам с максимальными радиусами. В 3-мерном пространстве (x, *y*, *z*), где *x* и *y* – это координаты точки на изображении, а *z* – величина большой полуоси эллипса с центром (*x*, *y*), соответствующего точке, локальные максимумы (рис. 7) будут отвечать наиболее вероятным положениям центров нот. На рис. 8 данные множества таких точек отображены белым цветом.

Как показало тестирование, одной ноте может соответствовать несколько областей локальных максимумов. Поэтому для уточнения положения центра каждой найденной ноты необходимо выполнить кластеризацию всех обнаруженных локальных максимумов, причём нормированная величина *z*-компоненты точки будет служить её весом.



Рисунок 8 – Наиболее вероятные центры нот

При кластеризации проводится поиск точек, удаленных от центра кластера не более, чем на длину малой полуоси *b*. Вычисление координаты нового центра кластера и его веса производится по следующей формуле

$$K = \left(K1 + K2\right) + \left(K1 - K2 \times \left(V2 - V1\right)\right)$$

$$V = \left(\left(K - K1\right) / \left(K2 - K1\right)\right) \times \left(V2 - V1\right) + V1, где$$

$K -$ *новая координата,*

$V -$ *новая вероятность,*

$K1 -$ *координата первой точки,*

$K2 -$ *координата второй точки,*

$V1 -$ *вероятность первой точки,*

$V2 -$ *вероятностьвторойточки.*

После кластеризации получаем искомые центры нот, на рис. 9 эти центры обозначены белым цветом.



Рисунок 9 – Результат кластеризации

Выводы

В ходе проведения этапа структурного анализа нотного изображения и детектировании нот на нем, был разработан метод детектирования нот, обладающий робастностью по отношению к качеству изображению нотного текста.

Робастность достигается за счет предварительной обработки, что позволило сделать следующие выводы.

Приведение изображения к оттенкам серого и использование порога, полученного экспериментально, позволяет отделить фон от значимых элементов даже на изображении плохого качества. При этом мелкие значимые части, которые в ходе предварительной обработки могут быть приняты за шум, не удаляются в результате структурного анализа.

Структурный анализ проводится на основе преобразования Хафа с учетом специфики нотного текста, имеющего иерархическую структуру. В результате с высокой точностью выделяются элементы каждого уровня иерархии, кроме нот.

Процедура детектирования нот базируется на геометрических свойствах искомых объектов. В ходе выполнения этой процедуры вычисляются радиусы эллиптических масок и за счет кластеризации уточняются центры эллипсов, что позволяет повысить эффективность детектирования нот по сравнению с методом поиска по маске, широко используемым при обработке цифровых изображений.

Продолжением данной работы является исследование эффективности методов распознавания выделенных элементов с целью создания технологии оптического распознавания нотного текста, позволяющей сделать процесс распознавания более автоматизированным и робастным по сравнению с существующими аналогами.

Литература

1. Шлезінгер М.І. Комп'ютерна технологія для розпізнавання типографських нотних текстів / М.І. Шлезінгер, Б.Д. Савчинський // Електронні зображення та візуальні мистецтва : зб. пр. Першої укр. конф. сер. EVA, 22-24 трав. 2002 р., Київ. – К., 2002. – С. 82-86. – Бібліогр. : 6 назв. – укр.
2. Дегтярева А. Преобразование Хафа / А Дегтярева, В. Вежневец [Электронный ресурс]. – Режим доступа : http://cgm.computergraphics.ru/content/view/36
3. Ильин В.А. Аналитическая геометрия / В.А. Ильин, Э.Г. Позняк. – Москва : Наука, 1988. – 223 с.
4. Кудрявцев Л.Д. Курс математического анализа / Кудрявцев Л.Д. – Москва : «Высшая школа», 1998. – 712 с.
5. Прэтт У. Цифровая обработка изображений : кн. 2 / Прэтт У. – М. :Мир, 1982. – 784 с.
6. Павлидис Т. Алгоритмы машинной графики и обработки изображений / Павлидис Т. – М. : Радио и связь, 1986. – 399 с.
7. Компьютерная система ввода и воспроизведения нотного текста <SOLO> // Искусственный интеллект. – 2005. – № 4. – С. 602-612.

8. Beran T. Recognition of printed music score / T. Beran, T. Macek // In : MLDM'99, NAI 1715, Springer-Verlag 1999. – P. 174-179.
9. Азаренко Д.С. Способ структурного анализа изображения нотного текста для дальнейшего распознавания / Д.С. Азаренко, А.В. Жук // Искусственный интеллект. – 2007. – №1 – С. 110-122.

Literatura

1. Komp'juterna tehnologija dlja rozpiznavannja typografs'kyh notnyh tekstiv. Elektronni zobrazhennja ta vizual'ni mistectva: Zb. pr. pershoi ukr. konf. ser. EVA. 22-24 trav. 2002 r. Kyiv. K. 2002. S. 82-86
2. Degtjareva A. Preobrazovanie Hafa. http://cgm.computergraphics.ru/content/view/36
3. Il'in V.A. Analiticheskaja geometrija. Moskva: Nauka. 1988. 223 s.
4. Kudrjavcev L.D. Kurs matematicheskogo analiza. Moskva: Vysshaja shkola. 1998. 712s.
5. Prjett U. Cifrovaja obrabotka izobrazhenij. Kn.2. M.:Mir. 1982. 784 s.
6. Pavlidis T. Algoritmy mashinnoj grafiki i obrabotki izobrazhenij. M.: Radio i svjaz'. 1986. 399 s.
7. Komp'juternaja sistema vvoda i vosproizvedenija notnogo teksta <SOLO>. Iskusstvennyj intellect. 2005. № 4. S. 602-612.
8. Beran T. In: MLDM'99. NAI 1715. Springer-Verlag. 1999. P. 174-179.
9. Azarenko D.S. Iskusstvennyj intellekt. 2007. №. 1. S. 110-122.

RESUME

D.S. Azarenko

Detection of Scores to the Note Text Image Based on its Structural Analysis

The article is devoted to the structural analysis of the note text image. The analysis includes preprocessing, obtaining a hierarchical structure of image text and detection notes on it.

As a result of the image preprocessing, relevant elements are separated from the background. For this purpose, the image is reduced to greyscale and a color histogram is constructed. Threshold for background separation is calculated on the basis of this histogram. Threshold is computed by the author with the formula which was obtained experimentally.

Structural analysis of the note text image is extraction of the elements of hierarchical structure. The Hough transform is used for structural analysis, this transform allows detecting lines in the image.

Notes are detected in staffs. Search of notes is performed by using templates. Template is elliptical mask, because the note is an ellipse which has a horizontal angle of inclination equal to 37 degrees.

On the basis of the proposed methods and approaches, optical recognition technology of note text can be developed. That allows making recognition process more automated and robust respect to the quality of the source data.

Статья поступила в редакцию 30.08.2012.